



An *in silico* approach to the analysis of CTLA-4 and CD28 molecules

Fuat Dilmeç¹, Lokman Varışlı², Osman Çen²

¹Harran University, Faculty of Medicine, Department of Medical Biology

²Harran University, Arts and Sciences Faculty, Department of Biology

(* author for correspondence; lokmanv@gmail.com)

Received 17 October 2006; Accepted 27 March 2007

Abstract

The CD28 and CTLA-4 are important co-stimulatory molecules expressed on T cells and regulate T cell activation through antigen stimulation. This study aimed to investigate some properties of these genes and its products; such as the homologous protein sequences to human CTLA-4 and CD28 molecules, the common transcription factor binding sites on their promoters, and phylogenetic relationships of their Immunoglobulin V-Type (IGv) domains using bioinformatical approaches. Our results revealed that both CTLA-4 and CD28 molecules are moderately conserved among all organisms investigated. The IGv domains of human CTLA-4 and CD28 are closest to those of *Macaca nemestrina*, *Cercocebus torquatus atys*, *Macaca mulatta* and *Papio cynocephalus anubis*. In contrast, the IGv domain of CD28 of *Gallus gallus* and that of CTLA-4 of *Mus musculus* and *Rattus norvegicus* is most apart from those of human, respectively. The comparative screening of the promoters revealed that both CTLA-4 and CD28 genes have some conserved transcription factor binding sites.

Key Words: CTLA-4, CD28, bioinformatics, comparative genomics, *in silico* biology

CTLA-4 ve CD28 moleküllerinin analizine *in silico* bir yaklaşım

Özet

CD28 ve CTLA-4, T hücreleri üzerinde eksprese olan önemli ko-stimilatör moleküllerdir ve antijen uyarımı ile T hücre aktivasyonunu düzenlerler. Bu çalışmanın amacı; omurgalı türlerin ilgili moleküllerinin insan CTLA-4 ve CD28 moleküllerine homolog protein dizilerini, promoter dizileri üzerinde genel transkripsiyon faktörlerinin bağlanma yerlerini ve farklı türlerdeki CTLA-4 ve CD28 proteinlerinin IGv domeynlerinin filogenetik ilişkileri gibi bu gen ve ürünlerinin özelliklerini biyoinformatik araçlarla araştırmaktır. Sonuçlarımız, hem CTLA-4 ve hem de CD28 moleküllerinin araştırılan tüm organizmalar arasında orta derecede korunduklarını gösterdi. Çalışmamız insanın hem CD28 hem de CTLA-4 IGv domeynlerinin; *Macaca nemestrina*, *Cercocebus torquatus atys*, *Macaca mulatta* ve *Papio cynocephalus anubis* 'in IGv domeynlerine evrimsel olarak en yakın olduğunu göstermiştir. Buna karşın, *Gallus gallus*'un CD28 ve *Mus musculus* ile *Rattus norvegicus*'un CTLA-4 IGv domeynleri ise sırasıyla insanlarınkine en uzaktır. Karşılaştırmalı promoter taraması, hem CTLA-4 ve hem de CD28 genlerinin evrimsel olarak çalışılan bütün türlerde korunmuş bazı transkripsiyon faktörlerinin bağlanma bölgelerine sahip olduklarını göstermiştir.

Anahtar Sözcükler: CTLA-4, CD28, biyoinformatik, karşılaştırmalı genomik, *in silico* biyoloji

Introduction

The CTLA-4 and CD28 molecules are expressed on T cells and interacts with the members of the B7 family of molecules expressed on antigen presenting cells (Dariavach et al., 1988; Harper et al., 1991). CD28 provides a co-stimulatory signal to the T cells leading to their activation, while CTLA-4 provides a negative signal and thus is thought to be an important regulator

of autoimmunity (King et al., 2002; Zhang et al., 2003).

CD28 and CTLA-4 molecules interact with their ligands through their immunoglobulin V-Type (IGv) domains. The IGv domains form a subgroup of immunoglobulin like domains. The members of this immunoglobulin like domains are involved in a variety of functional structures, including cell-surface receptors, muscle structure, and the immunological

Table 1. The homolog clones to human CD28 and CTLA-4 molecules as obtained by BLASTp. A higher similarity may indicate more conserved and closer evolutionary relationship.

	Species	Common name	AC No	% Identity with human CD28
CD28	Homo sapiens	Human	NP006130	100
	Macaca nemestrina	Pigtailed monkey	AAK37612	97
	Cercocebus torquatus atys	White collared (sooty) mangabey	AAK37601	97
	Papio cynocephalus anubis	Olive baboon	AAK37533	96
	Macaca mulatto	Rhesus monkey	AAK37538	95
	Callithrix jacchus	Common marmoset	AAK37604	85
	Canis familiaris	Dog	AAG43370	80
	Felis catus	Feral cat	BAA92349	79
	Oryctolagus cuniculus	European rabbit	BAA08641	78
	Marmota monax	Woodchuck	AAF36501	76
	Bos taurus	Cow	CAA63707	77
	Ovis aries	Sheep	AAD04379	77
	Rattus norvegicus	Rat	CAA39003	69
	Mus musculus	Mouse	AAH64058	69
	Gallus gallus	Chicken	CAA48114	50
	Species	Common name	AC No	% Identity with human CTLA-4
CTLA-4	Homo sapiens	Human	NP005205	100
	Papio cynocephalus anubis	Olive baboon	AAK37534	97
	Macaca nemestrina	Pigtailed monkey	AAK37537	96
	Cercocebus torquatus atys	White collared (sooty) monkey	AAK37608	96
	Macaca mulatto	Rhesus monkey	AAK37605	96
	Aotus trivirgatus	Owl monkey	AAK37530	91
	Felis catus	Feral cat	AAD50988	87
	Canis familiaris	Dog	AAF23813	87
	Sus scrofa	Pig	AAF86981	85
	Oryctolagus cuniculus	European rabbit	BAA08644	84
	Marmota monax	Woodchuck	AAF36502	86
	Bos Taurus	Cow	CAA63708	84
	Ovis aries	sheep	AAD04380	83
	Rattus norvegicus	Rat	AAC52502	75
	Mus musculus	Mouse	AAF01489	74

[illegible][illegible]

synapse components (Teichmann and Chothia, 2000). IGv domains are found in diverse protein families, including immunoglobulin light and heavy chains; T-cell receptors, membrane and junctional adhesion molecules; tyrosine-protein kinase receptors; and in some of the apoptotic surface molecules (Smith and Xue, 1997).

conservation among these genes in human, mouse and rat (Martin et al., 2000, Wang et al., 2006). Gross and his colleagues (1990) had shown that the murine CD28 clone had 61% nucleotide identity with the human CD28 cDNA and both the murine and human CD28 molecules were integral membrane glycoproteins with a hydrophobic signal peptide sequence and a transmembrane region. (Gross et al., 1990).

In this study, we aimed to investigate the phylogenetic relationship of IGv domains, the transcription factor binding sites on promoters, and protein homology of human CTLA-4 and CD28 to those of other vertebrates using bioinformatics tools.

We used Genomatrix software (<http://www.genomatrix.de>) for analysis of CTLA-4 and CD28 gene pro-

B.

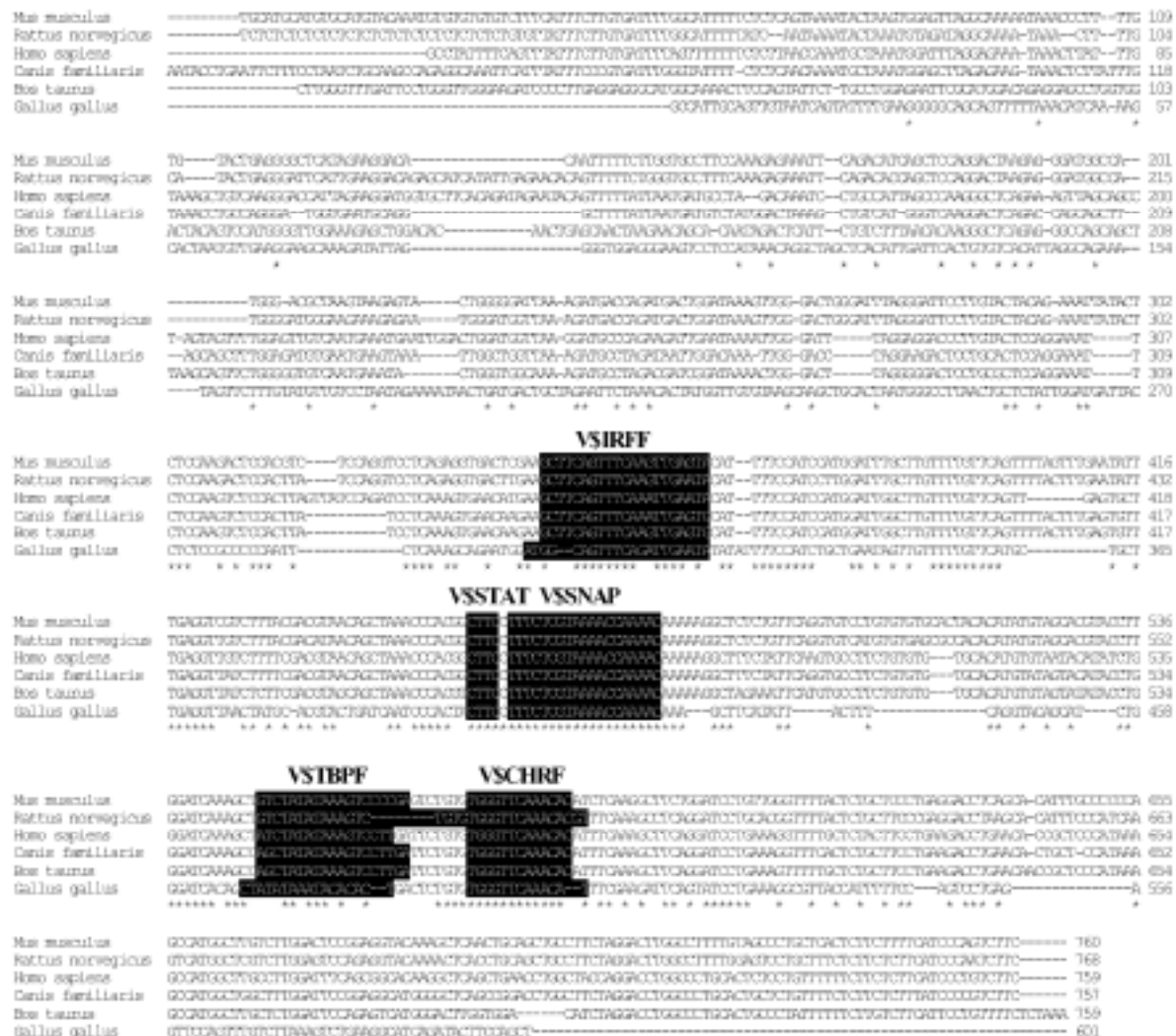


Figure 2. Multiple alignments of CD28 (A) and CTLA-4 (B) gene promoters present in Genomatix database. Nucleotide positions are indicated on the right side. Nucleotides conserved among these species are shown by an asterisk under the alignment. Conserved elements are shaded with black background. The terms used in the figure and their explanations are as following: VSIRFF-Interferon regulatory factor, V\$STAT-Signal transducer and activator of T cells, V\$SNAP-Proximal sequence element of RNA polymerase III-transcribed genes, V\$TBPF-Avian C-type LTR TATA box, V\$CHRF-Cell cycle gene homology region, V\$HAML-Human acute myelogenous leukemia factors.

motors in various species. In the Genomatix database, there are six CTLA-4 and CD28 promoters, those of *H. sapiens*, *B. taurus*, *C. familiaris*, *R. norvegicus*, *M. musculus* and *G. gallus*. These nucleotide sequences were downloaded and then were aligned using the

ClustalW program. Then, *in silico* common transcription factor binding sites were searched through the Dialign TF program in the Genomatix software for all the promoters present in the database.

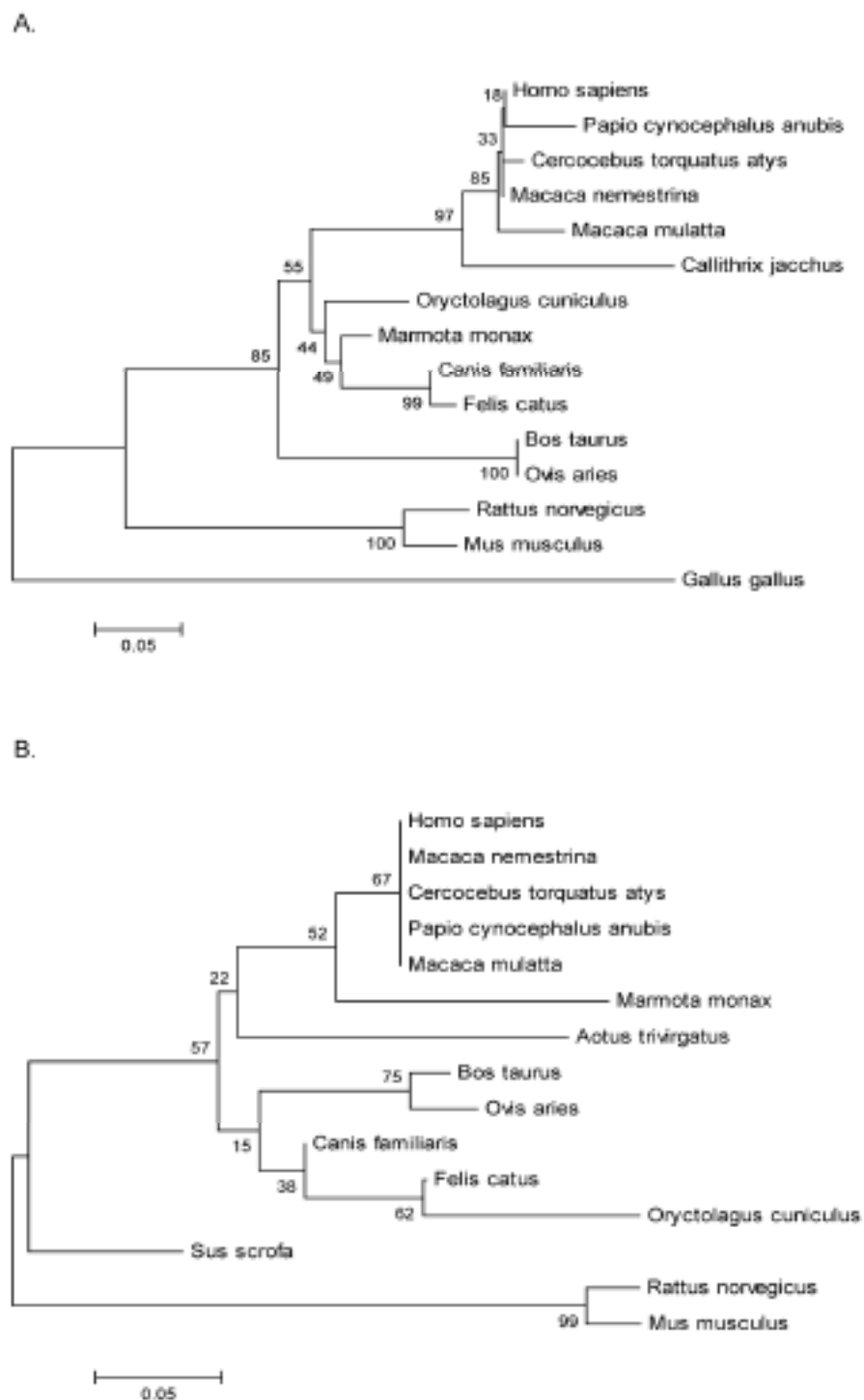


Figure 3. Phylogenetic trees of CD28 (A) and CTLA-4 (B) IGv domains. Phylogenetic trees were constructed by MEGA 3. Species names and branch lengths are indicated. Branch lengths indicate evolutionary relationship. A. Phylogenetic tree of CD28 IGv domains in various species and B. Phylogenetic tree of CTLA-4 IGv domains in various species.

Evolutionary analysis

We used amino acid sequences of IGv domains of both CD28 and CTLA-4 molecules to construct phylogenetic trees using the neighbor-joining method (NJ) with Jones-Taylor-Thomton (JTT) distances. NJ searches were conducted using MEGA 3 program (Kumar et al., 2004) and 500 bootstrap replicates were assessed for the reliability of internal branches; sites with gaps were ignored in the analysis.

Results

Homology search

BLASTP of human CD28 and CTLA-4 revealed that both CD28 and CTLA-4 are found in various vertebrate species (Table 1). When the proteins of human CD28 and CTLA-4 genes were compared with those of the other vertebrate species, it was seen that human CD28 protein showed the highest similarity to that of *M. nemestrina* (97%), whereas CTLA-4 protein to that of *P. cynocephalus Anubis* (97%). In contrast, human CD28 and CTLA-4 proteins showed the lowest similarities to those of *G. gallus* (50%) and *M. musculus* (74%), respectively (Table 1).

ClustalW alignment of IGv domains of both CD28 and CTLA-4 molecules revealed that these protein domains are well conserved especially in some localized areas (Figure 1). “TDIYFCK” and “VNLTIQGL” motifs in the IGv domains of CD28 and CTLA-4, respectively, are found to be well conserved in all investigated species (Figure 1). These regions are most probably functionally important and any mutations in them are deleterious, as implicated by their evolutionary conservation. Using the NetNGlyc program, we predicted a putative N-glycosylation site in the conserved “VNLTIQGL” region.

Promoter analysis

By database searches we found that while the orthologous CTLA-4 gene promoters include five (V\$IRFF, V\$STAT, V\$SNAP, V\$TBPF and V\$CHRF) common transcription factor binding sites (TFBs), the orthologous CD28 gene promoters harbor only V\$HAML common transcription factor binding site (Figure 2). The family of V\$IRFF comprises six (IRF1, IRF2, IRF3, IRF4, IRF7 and ISRE) different transcription factor binding sites.

Evolutionary analysis

From the phylogenetic trees constructed by the MEGA 3 (Figure 3), we found that both CTLA4 and CD28 IGv domains were conserved among all organisms investigated.

Discussion

CD28 and CTLA-4 co-stimulatory molecules expressed on T lymphocytes and interact with their ligands B7-1 (CD80) and B7-2 (CD86) with their IGv domains along with antigenic stimulation of T cells through the T cell receptor-antigen-MHC (TCR-antigen-MHC) engagement. This interaction provides T cells with a second signal leading to their activation. Without this second signal, T cells, stimulated through TCR-antigen-MHC, will not get activated and undergo apoptosis if they do not receive a second signal thorough these co-stimulatory molecules. Therefore, for a successful immunological response, these molecules hold an important role by generating a required second signal for the T cell activation.

Several researchers have shown that there was 100% conservation in the amino acid sequences of the cytoplasmic tails of human, mouse, and rat CTLA-4 molecules (Martin et al., 2000; Harper et al., 1991). Likewise, in the present study we showed that the BLASTp results of human CD28 and CTLA-4 revealed that both CD28 and CTLA-4 are found in various species of vertebrates and these molecules have 50-97% conservation degree in the total amino acid sequences (Table 1). The human CD28 molecule has highest homology to those of pigtailed monkey and sooty mangabey (97%) and lowest homology to that of chicken (50%). On the other hand, we found that the human CTLA-4 shows the highest homology to that of olive baboon (97%) and the lowest homology to that of mouse (74%) (Table 1). These results suggest that the CTLA-4 and CD28 genes have well been conserved during evolution.

We also examined the phylogenetic trees of IGv domains of CTLA-4 and CD28 in various species using MEGA 3 for Evolutionary Analysis (Figure 3). We observed that (Figure 1 and 2) human IGv domains of CD28 and CTLA-4 proteins show closest homology to those of pigtailed monkey, sooty mangabey, rhesus monkey, and olive baboon. Likewise, same figures show that IGv domains of CD28 of chicken and CTLA-4 of mouse and rat are most diverse to IGv

domain of human CD28 and CTLA-4 molecules, respectively. Closest homology of pigtailed monkey, rhesus monkey, sooty mangabey, and olive baboon sequences to human sequences is consistent with the general evolutionary data related to these organisms, all of which are primates. All of the primates investigated in the study belong to old world monkey group, except human. According to the general evolutionary data macaque monkeys are closest organisms to human among the old world monkeys. On the general phylogenetic tree, after macaque monkeys baboons and finally mangabeys come. However, when we examined phylogenetic tree and multiple alignment of IGv domains of CD28 molecules, we see that all the primates have the same amino acid sequence, and as a consequence, there are not any differences in their domains. Therefore, on the basis of these IGv domains, these organisms do not display any differences. Likewise, when we examined phylogenetic tree of IGv domains of CTLA-4 molecules, we see that location of species on the tree isn't consistent with general evolutionary data.

When we compared the conserved IGv domains among different species, we found "TDIYFCK" and "VNLTIQGL" sequence motifs on CD28 and CTLA-4, respectively. "MYPPPY" motif mediates binding to the ligands (B7-1/B7-2) and is located outside the conserved domain. "TDIYFCK" and "VNLTIQGL" motifs we determined are likely to bind various antibodies. But it requires further experimental investigations for support.

We also performed a posttranslational modification and functional motif searches to this conserved regions. The results of these analysis revealed that conserved "VNLTIQGL" region has an N-Glycosylation motif. N-glycosylation is a common posttranslational modification to the membrane proteins. This motif has N-{P}-[ST]-{P} pattern, in which a proline in the second or forth position after the modified N is forbidden and a serine or threonine is required in the third position. The "TDIYFCK" region found in the CD28 IGv domain did not seem to harbor any known motifs. However, it is possible that it may undergo phosphorylation on either treonine or tyrosine residues. It had been shown that the antibodies were focused on a region around phenylalanine residue 98, which lies adjacent to the highly conserved "MYPPPY" loop found in both CD28 and CTLA-4 (Zhang et al., 2003). In the same way, in cancer therapy, differ-

ent CTLA-4 and specific antibodies have been used and consequently the human lymphocyte mediated tumor suppression was significantly enhanced in mice co-engrafted with tumor cells (Sabel et al., 2005).

We used Dialign TF program in Genomatix software to predict transcription factor binding sites (transcriptional elements) of all orthologous CTLA-4 and CD28 promoters present in the database. Dialign TF results revealed that both CD28 and CTLA-4 orthologous promoters have common transcriptional elements. The conservation of transcriptional elements in promoter sequences is the result of non-redundant functional importance (Cartharius et al., 2005). This theory has been recently proven by different investigators studying CTNNA3 and Gfi1 genes promoters (Vanpoucke et al., 2004; Doan et al., 2004).

Basic bioinformatic techniques are powerful tools in terms of discovering and analyzing novel genes (Hu et al., 2005). Recently, we identified and characterized two novel genes using bioinformatics tools (Varisli and Cen, 2005; Varisli and Cen, 2007). Even though the results from bioinformatical studies are very helpful in directing and designing the experiments, they principally need to be experimentally confirmed.

Acknowledgements

This work was partially supported by Harran University Research and Development Institute to L.V. (graduate student grant no: 432).

References

- Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 215: 403-410, 1990.
- Blom N, Gammeltoft S and Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol.* 294(5): 1351-1362, 1999.
- Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M and Werner T. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics.* 21: 2933-2942, 2005.
- Dariavach P, Mattei MG, Golstein P and Lefranc MP. Human Ig superfamily CTLA-4 gene: chromosomal localization and identity of protein sequence between murine and human CTLA-4 cytoplasmic domains. *Eur J Immunol.* 18(12): 1901-1905, 1988.
- Doan LL, Porter SD, Duan Z, Flubacher MM, Montoya D, Tschlis PN, Horwitz M, Gilks CB and Grimes HL.

- Targeted transcriptional repression of Gfi1 by GFI1 and GFI1B in lymphoid cells. *Nucleic Acids Res.* 32, 2508-2519, 2004.
- Duckert P, Brunak S and Blom N. Prediction of proprotein convertase cleavage sites. *Protein Eng Des Sel.* 17, 107-112, 2004.
- Gross JA, St John T and Allison JP. The murine homologue of the T lymphocyte antigen CD28. Molecular cloning and cell surface expression. *J Immunol.* 15: 144(8): 3201-3210, 1990.
- Harper K, Balzano C, Rouvier E, Mattei MG, Luciani MF and Golstein P. CTLA-4 and CD28 activated lymphocyte molecules are closely related in both mouse and human as to sequence, message expression, gene structure, and chromosomal location. *J Immunol.* 1: 147(3): 1037-1044, 1991.
- Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG and Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680, 1994.
- Hu Z, Chen K, Wang L, and Yao Q. Identification and characterization of *Bombyx mori* eIF5A gene through bioinformatics approaches. *In Silico Biol.* 5: 573-580, 2005.
- Julenius K, Molgaard A, Gupta R and Brunak S. Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology.* 15(2):153-64, 2005.
- King AL, Moodie SJ, Fraser JS, Curtis D, Reid E, Dearlove AM, Ellis HJ and Ciclitira PJ. CTLA4/CD28 gene region is associated with genetic susceptibility to coeliac disease in UK families *J Med Genet.* 39: 51-54, 2002.
- Kumar S, Tamura K and Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinf.* 5: 150-163, 2004.
- Martin KO, Karen MH, Penwell RT, Stauber EC, Warren SJ and Tector AJ. A native soluble form of CTLA-4. *Cellular Immunology.* 201, 144-153, 2000.
- Sabel MS, Hess SD, Egilmez NK, Conway TF Jr, Chen FA, Bankert RB. CTLA-4 blockade augments human T lymphocyte-mediated suppression of lung tumor xenografts in SCID mice. *Cancer Immunol Immunother.* 54(10): 944-952, 2005.
- Smith DK and Xue H. Sequence profiles of immunoglobulin and immunoglobulin-like domains. *J Mol Biol.* 274: 530-545, 1997.
- Teichmann SA and Chothia C. Immunoglobulin superfamily proteins in *Caenorhabditis elegans*. *J Mol Biol.* 296: 1367-1383, 2000.
- Vanpoucke G, Goossens S, De Craene B, Gilbert B, van Roy F and Berx G. GATA-4 and MEF2C transcription factors control the tissue specific expression of the alpha T-catenin gene *CTNNA3* *Nucleic Acids Res.* 32, 4155-4165, 2004.
- Varisli L and Cen O. Identification and characterization of Rat GMDS gene by using bioinformatics tools. *Turk J Biochem.* 30: 306-309, 2005.
- Varisli L and Cen O. Identification and characterization of rat ETV6 gene by using bioinformatics tools. *JABS.* 1: 19-23, 2007.
- Wang X, Ma Z, Hong Y, Lu P and Lin Q. Expression of CD28 and cytotoxic T lymphocyte antigen 4 at the maternal-fetal interface in women with unexplained pregnancy loss. *Int J Gynaecol Obstet.* 93(2): 123-129, 2006.
- Waterson RH, Lindblad-Toh K and Birney E. Initial sequencing and comparative analysis of the mouse genome. *Nature.* 420: 560-562, 2002.
- Zhang X, Schwartz JDC, Almo SC and Nathenson SG. Crystal structure of the receptor-binding domain of human B7-2: Insights into organization and signaling. *PNAS* 100(5): 2586-2591, 2003.